

**Título de la tarea: Prediseño de un sistema Big Data**  
**Curso de Especialización: Inteligencia Artificial y Big Data**  
**Módulo profesional: Sistemas de Big Data**

**¿Qué contenidos o resultados de aprendizaje trabajaremos?**

**Resultados de aprendizaje**

**RA1.** Aplica técnicas de análisis de datos que integran, procesan y analizan la información, adaptando e implementando sistemas que las utilicen.

**Contenidos:**

- 1.- Por qué Big Data.
  - 1.1.- Las 5 Vs.
    - 1.1.1.- Volumen.
    - 1.1.2.- Velocidad.
    - 1.1.3.- Variedad.
    - 1.1.4.- Veracidad.
    - 1.1.5.- Valor.
  - 1.2.- Qué conseguimos gracias a Big Data.
    - 1.2.1.- Desde los eventos al valor.
- 2.- Clusters de computadoras.
- 3.- Conceptos de almacenamiento de datos.
  - 3.1.- Dataset.
  - 3.2.- Almacén de Datos.
  - 3.3.- ACID.
  - 3.4.- Teorema CAP.
  - 3.5.- BASE.
- 4.- Conceptos de procesamiento de datos.
  - 4.1.- Procesamiento en paralelo.
  - 4.2.- Procesamiento distribuido.
  - 4.3.- Estrategias de procesamiento de datos.
  - 4.4.- OLTP.
  - 4.5.- OLAP.
  - 4.6.- Principio SCV.
- 5.- La arquitectura por capas de Big Data.
- 6.- El paisaje de Big Data.

## Descripción de la tarea.

### Caso práctico

La empresa **Construcciones D8** se ha puesto en contacto con la empresa consultora en la que trabajas para que les realicéis un prediseño de lo que sería un sistema Big Data para resolver las siguientes necesidades.

- Hay distintas fuentes externas a su empresa que producen datos interesantes para ellos y les interesaría poder conectarse a ellas para obtenerlos.
- Esas fuentes tienen conjuntos de datos estáticos o que se actualizan anualmente.
- Además hay fuentes internas de la propia empresa que generan datos de forma continua y hay que irlos obteniendo sobre la marcha.
- La cantidad de datos actualmente es de aproximadamente 500TB, y calculan que se producen otros 100TB nuevos cada año.
- Quieren poder mantener almacenados todos esos datos de modo no se pierdan y además accesibles en todo momento.
- Se realizan transacciones debido a la interacción con clientes en el día a día.
- La junta directiva se reúne una vez al mes y quiere poder acceder a un cuadro de mandos para ver analíticas descriptivas que empleen todos los datos que estuviesen disponibles una semana antes de reunirse. Tales analíticas deben ser interactivas, siendo los directivos capaces de realizar filtrados de información de modo que las gráficas mostradas se actualicen según la información seleccionada.
- Quieren poder decidir a qué clientes ofrecerles ciertas ofertas en función de lo que se sabe de su comportamiento pasado.

### ¿Qué te pedimos que hagas?

#### Apartado 1: Prediseña un sistema para Big Data

Crea un documento en el que explicas cómo sería el sistema a emplear para resolver las necesidades Big Data del supuesto práctico. Deberás:

- Indicar qué habrá que hacer para ir aumentando la capacidad del cluster según se reciben nuevos datos.
- Indicar qué capas de la arquitectura Big Data necesitarán estar presentes como mínimo en el sistema a crear.
- Indicar si alguna parte del sistema necesitará cumplir con las características ACID.
- Indicar si será necesario un subsistema OLTP.
- Indicar si será necesario un subsistema OLAP.
- Indicar si habrá un almacén de datos.

- Indicar qué estrategia de procesamiento habrá que emplear para poder crear el cuadro de mandos que quiere la junta directiva.
- Indicar si será necesario crear modelos predictivos a partir de los datos.

### **Información de interés.**

## **Recursos necesarios y recomendaciones**

### **Recursos necesarios**

Ordenador personal capaz de ejecutar un editor de textos.

Editor de textos.

### **Recomendaciones**

Antes de abordar la tarea:

Lee con detenimiento la unidad, consulta información en internet para saber más, examina el material proporcionado por el profesor y aclara las dudas que te surjan con él.

Realiza el cuestionario 1 de la unidad, y consulta nuevamente las dudas que te surjan.

Sólo cuando lo tengas todo claro, debes abordar la realización de la tarea.

No olvides elaborar el documento explicativo.

### **Indicaciones de entrega**

Una vez realizada la tarea, el envío se realizará a través de la plataforma. El archivo se nombrará siguiendo las siguientes pautas:

**Apellido1\_Apellido2\_Nombre\_SBD01\_Tarea**

Asegúrate que el nombre no contenga la letra ñ, tildes ni caracteres especiales extraños.

Así por ejemplo la alumna **Marta Sánchez Lopez para la primera unidad del MP de SBD**, debería nombrar esta tarea como...

**sanchez\_lopez\_marta\_SBD01\_Tarea**

## **Evaluación de la tarea.**

### **Criterios de evaluación implicados**

#### **Criterios de evaluación RA1**

e. Se han establecido objetivos y prioridades, secuenciación y organización del tiempo de realización.

f. Se han seleccionado e integrado sistemas de información que satisfacen las necesidades del problema.

g. Se han determinado criterios de coste y calidad necesarios para la eficacia y eficiencia de la implementación de un sistema Big Data.

## ¿Cómo valoramos y puntuamos tu tarea?

Rúbrica de la tarea	
<b>Apartado:</b> Indica correctamente qué habrá que hacer para ir aumentando la capacidad del cluster según se reciben nuevos datos.	1,25 puntos
<b>Apartado:</b> Indica correctamente qué capas de la arquitectura Big Data necesitarán estar presentes como mínimo en el sistema a crear.	1,25 puntos
<b>Apartado:</b> Indica correctamente si alguna parte del sistema necesitará cumplir con las características ACID.	1,25 puntos
<b>Apartado:</b> Indica correctamente si será necesario un subsistema OLTP.	1,25 puntos
<b>Apartado:</b> Indica correctamente si será necesario un subsistema OLAP.	1,25 puntos
<b>Apartado:</b> Indica correctamente si habrá un almacén de datos.	1,25 puntos
<b>Apartado:</b> Indica correctamente qué estrategia de procesamiento habrá que emplear para poder crear el cuadro de mandos que quiere la junta directiva.	1,25 puntos
<b>Apartado:</b> Indica correctamente si será necesario crear modelos predictivos a partir de los datos	1,25 puntos
Redacción clara y correcta, sin errores ortográficos	<b>Se resta 0,1 puntos por cada error ortográfico o expresiones incorrectas</b>